

ETUDE DES HYPOTHÈSES DE L'ANALYSE DE LA VARIANCE À DEUX CRITÈRES DE CLASSIFICATION : APPROCHE PAR L'EXEMPLE

A.F. IEMMA⁽¹⁾ et J.J. CLAUSTRIAUX⁽²⁾

RÉSUMÉ

A partir d'exemples simples et des résultats fournis par les procédures d'analyse de la variance des logiciels SAS et Minitab (GLM), on présente les hypothèses qui peuvent être construites pour analyser un tableau de données à deux dimensions.

SUMMARY

By examples and results from SAS and Minitab softwares (GLM), analysis of variance hypothesis built to analyse a two dimensional data table are introduced.

MOTS-CLEFS

Analysis of variance, non-orthogonal data, tests of hypothesis, statistical software.

1. INTRODUCTION

En vue de réaliser une analyse de la variance à deux critères de classification, modèle croisé fixe, les données peuvent être rassemblées dans un tableau à deux dimensions ($i = 1, \dots, p$; $j = 1, \dots, q$).

L'examen des effectifs n_{ij} des pq échantillons ou fréquences des pq combinaisons ou cellules ij , permet de qualifier le tableau par les termes complet, orthogonal et équilibré, définis dans une précédente publication [CLAUSTRIAUX et IEMMA, 1999].

-
- (1) Professeur à l'*Escola Superior de Agricultura Luiz de Queiroz*, Piracicaba, São Paulo (Brésil).
 - (2) Professeur à la Faculté universitaire des Sciences agronomiques de Gembloux (Belgique).

En pratique, en présence d'au moins une fréquence n_{ij} supérieure à 1, cinq types de tableaux peuvent être identifiés.

L'analyse des données de ces tableaux par la procédure GLM des logiciels SAS [X, 1990] et Minitab [X, 1996] aboutit respectivement à plusieurs propositions de résultats conduisant à une certaine confusion si, préalablement, la notion d'hypothèse à tester n'a pas été parfaitement définie, SAS proposant quatre types d'hypothèses et Minitab deux types d'hypothèses.

La figure 1 présente les types de tableaux et les identités ou non entre les hypothèses de SAS (caractères gras) et de Minitab (caractères gras italiques), en considérant chaque fois un modèle avec et sans interaction.

Pour éclairer ces différentes situations, trois cas sont étudiés :

- un tableau incomplet, non orthogonal et déséquilibré (paragraphe 3) : une ou plusieurs combinaisons ij ne sont pas représentées faisant apparaître au moins une cellule ij vide ou cas avec cellules vides ($n_{ij} \neq n$; $n_{ij} \geq 0$);
- un tableau complet, non orthogonal et déséquilibré (paragraphe 4) : chaque combinaison ij est répétée de façon relativement différente ou cas non équiréparté ($n_{ij} \neq n$; $n_{ij} > 0$);
- un tableau complet, orthogonal et équilibré (paragraphe 5) : chaque combinaison ij des deux critères de classification est répétée un même nombre de fois ou cas équiréparté ($n_{ij} = n$; $n_{ij} > 0$).

Ces trois cas sont illustrés par les données (dosages de P_2O_5 en mg pour 100 g de terre sèche) issues d'une étude destinée à comparer dans deux types de sol (facteur A; $i=1, 2$) trois sondes pédologiques (facteur B; $j = 1, 2, 3$) permettant de prélever des échantillons de terre ($k = 1, 2, 3, 4$) à une même profondeur [DAGNELIE, 1998].

Le tableau 1 rassemble l'ensemble des données illustrant le cas équiréparté. Si on suppose que les données en caractères italiques ($y_{124} = 44$; $y_{134} = 48$) n'ont pas été observées, on aboutit au cas non équiréparté. Si les données mises en caractères gras sont seulement considérées, le tableau résultant correspond au cas avec cellules vides.

Préalablement, pour mieux saisir les considérations appliquées, il importe de présenter quelques principes et notations plus théoriques (paragraphe 2). En guise de conclusion, on attire, notamment, l'attention du lecteur sur les dangers d'utiliser l'analyse de la variance sans discernement et sans réflexion préalable sur les hypothèses (paragraphe 6).

Deux remarques doivent encore être formulées :

- les techniques d'estimation des données manquantes ne sont pas abordées d'autant plus que des problèmes subsistent lorsque le nombre de données manquantes est élevé [FREUND, 1980; DODGE, 1985; MURRAY, 1986];
- le cas particulier d'un tableau à effectifs proportionnels [CLAUSTRIAUX, IEMMA, 1999; URQUARDT, WEEKS, 1978] n'est pas considéré, sachant que certains auteurs en contestent les avantages [BURDICK, HERR, 1980].

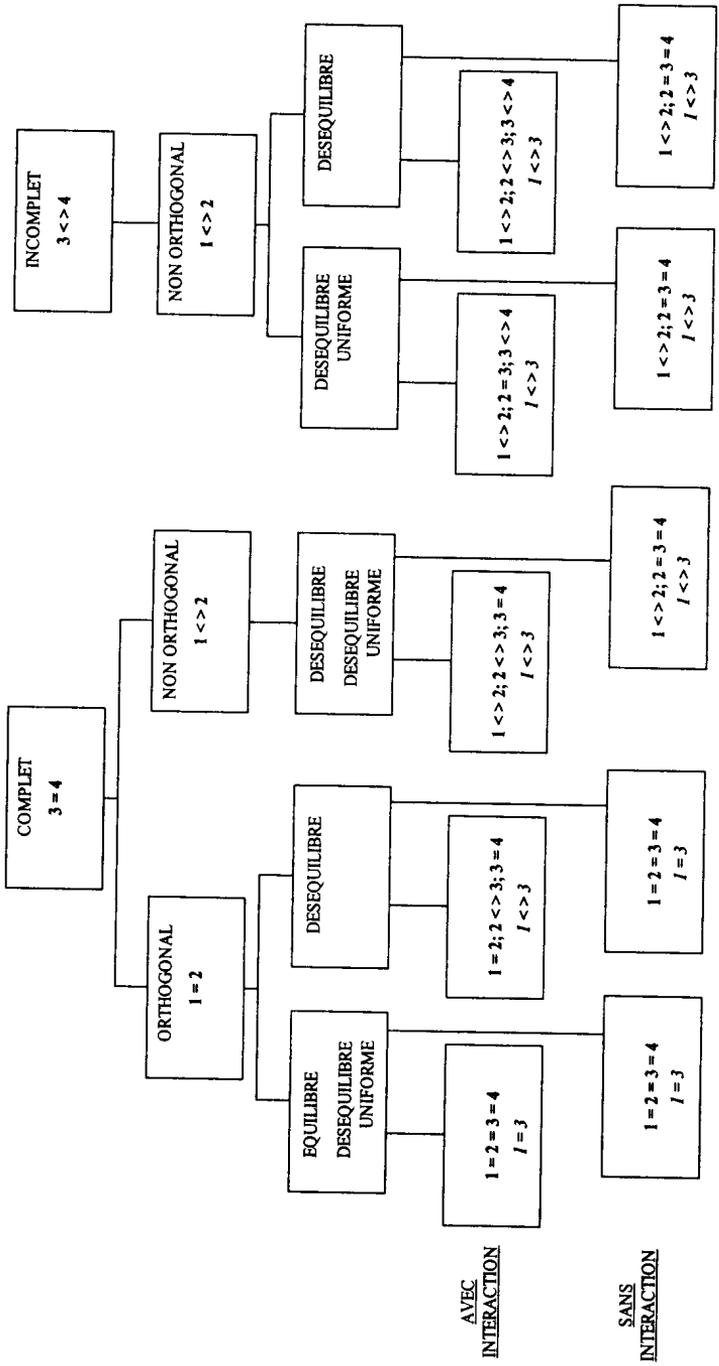


Figure 1. Types de tableaux et d'hypothèses SAS et Minitab associés.

Tableau 1. Tableau complet des données.

	Sonde 1 (j=1)	Sonde 2 (j=2)	Sonde 3 (j=3)	Moyennes ($\bar{y}_{i..}$)
Sol 1 (i=1)	43	41	42	44,75
	45	42	44	
	46	43	46	
	53	44	48	
	_____	_____	_____	
	46,75	42,50	45,00	
Sol 2 (i=2)	40	35	37	39,25
	40	37	39	
	40	40	40	
	43	40	40	
	_____	_____	_____	
	40,75	38,00	39,00	
Moyennes ($\bar{y}_{.j}$)	43,75	40,25	42,00	42,00 ($\bar{y}_{..}$)

Signalons aussi que des publications [HERR, 1986; YATES, 1934] présentent ces notions souvent sous des aspects théoriques, laissant finalement le praticien perplexe, même lorsqu'il est fait référence à l'emploi de certains logiciels [AZAIS, 1994].

Enfin, pour les exemples traités, les conditions d'application de l'analyse de la variance sont supposées être respectées et le niveau de signification choisi pour discuter les résultats est fixé à 5 %. Dans les figures, toutes les informations présentées en italique ne font pas partie des résultats proposés par les logiciels SAS et Minitab. La mise en évidence en caractères gras de certaines rubriques sert uniquement à attirer l'attention du lecteur par référence à des éléments repris dans le texte de la publication.

2. PRINCIPES THÉORIQUES

En général, l'analyse de la variance se fonde sur le modèle linéaire de Gauss-Markov:

$$y = X\theta + \epsilon, \quad (1)$$

où y est un vecteur des réalisations de la variable aléatoire Y (dosages), X est une matrice des constantes identifiant le plan expérimental, de rang complet ou incomplet, θ est un vecteur de paramètres inconnus en relation directe avec les hypothèses étudiées et ϵ est un vecteur des variations aléatoires non observables, supposées être distribuées selon des distributions normales de moyennes nulles et de variances constantes σ^2 .

1° Ce modèle (1) peut être présenté de plusieurs façons selon la structure choisie pour le vecteur θ . En relation avec l'analyse de la variance, différents modèles sont suggérés pour les réalisations y_{ijk} du vecteur y , chaque fois accompagnés d'une notation R suivie entre parenthèses du type de décomposition du vecteur θ :

- toutes les réalisations sont décrites par un seul paramètre μ :

$$y_{ijk} = \mu + \epsilon_{ijk}^{(1)} \quad (2)$$

et

$$R(\mu);$$

- à la constante μ s'ajoute l'effet des sols α_i ou des sondes β_j :

$$y_{ijk} = \mu + \alpha_i + \epsilon_{ijk}^{(2)} \quad (3)$$

et

$$R(\mu, \alpha);$$

ou

$$y_{ijk} = \mu + \beta_j + \epsilon_{ijk}^{(3)} \quad (4)$$

et

$$R(\mu, \beta);$$

- à la constante μ s'ajoute les deux effets α_i et β_j :

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk}^{(4)} \quad (5)$$

ou

$$y_{ijk} = \mu + \beta_j + \alpha_i + \epsilon_{ijk}^{(4)} \quad (5)$$

et

$$R(\mu, \alpha, \beta);$$

- à la constante μ s'ajoute, non seulement, les deux effets, mais aussi leur interaction :

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}^{(5)} \quad (6)$$

ou

$$y_{ijk} = \mu + \beta_j + \alpha_i + \gamma_{ij} + \varepsilon_{ijk}^{(5)} \quad (6)$$

et

$$R(\mu, \alpha, \beta, \gamma).$$

Il faut remarquer que les résidus des modèles $\varepsilon_{ijk}^{(1)}$ à $\varepsilon_{ijk}^{(6)}$ sont tous différents.

2° A chaque modèle correspond une équation des sommes des carrés (SC) :

$$SC_{\text{total}} = SC_{\text{paramètres}} + SC_{\text{résidus}}$$

auxquelles les degrés de libertés (d.l.) sont associés :

$$d.l._{\text{total}} = d.l._{\text{paramètres}} + d.l._{\text{résidus}}.$$

3° Il y a une analogie entre les notations R et $SC_{\text{paramètres}}$. Par exemple, pour obtenir uniquement l'effet des sols, noté $R(\alpha|\mu)$, ignorant totalement l'effet des sondes, ou inversement, c'est-à-dire la somme des carrés de A "ajustée" par la moyenne générale, il suffit d'écrire :

$$R(\mu, \alpha) - R(\mu) = R(\alpha|\mu),$$

ou

$$R(\mu, \beta) - R(\mu) = R(\beta|\mu).$$

Dès à présent, signalons que les deux sommes des carrés (SC) correspondant à $R(\alpha|\mu)$ et $R(\alpha|\beta)$ sont celles proposées par le logiciel SAS (type 1) pour les facteurs A et B (paragraphes 3.2 et 3.3).

De façon analogue, pour obtenir la somme des carrés de A ajustée par B (sols ajustés par les sondes) ou celle de B ajustée par A (sondes ajustées par les sols), il faut écrire :

$$R(\mu, \alpha, \beta) - R(\mu, \beta) = R(\alpha|\mu, \beta),$$

ou

$$R(\mu, \alpha, \beta) - R(\mu, \alpha) = R(\beta|\mu, \alpha).$$

Les sommes des carrés de ces deux cas ($R(\alpha|\mu, \beta)$ et $R(\beta|\mu, \alpha)$) sont celles du type 2 de SAS pour A et B (paragraphes 3.2 et 3.3).

Enfin, pour la somme des carrés de l'interaction, l'équation est :

$$R(\mu, \alpha, \beta, \gamma) - R(\mu, \alpha, \beta) = R(\gamma|\mu, \alpha, \beta).$$

Les quatre types de SAS proposent la somme des carrés de l'interaction correspondante; l'hypothèse de l'interaction peut donc être testée par l'un quelconque des quatre types (paragraphe 3.4).

Remarquons encore que si on choisit le modèle (6) en introduisant les restrictions suivantes :

$$\sum_{i=1}^p \alpha_i = 0,$$

$$\sum_{j=1}^q \beta_j = 0,$$

$$\sum_{i=1}^p \gamma_{ij} = 0, \quad \forall j \quad \text{et} \quad \sum_{j=1}^q \gamma_{ij} = 0, \quad \forall i,$$

on obtient les sommes des carrés $R(\alpha|\mu, \beta, \gamma)$ et $R(\beta|\mu, \alpha, \gamma)$ par le type 3 de SAS (paragraphe 3.4).

4° Le tableau d'analyse de la variance complet (tableau 2), établi pour le cas équilibré (tableau 1), permet d'illustrer la notation R et l'analogie avec les sommes des carrés des paramètres. Comme le tableau est orthogonal, les égalités suivantes sont établies [CLAUSTRIAUX, IEMMA, 1999] :

$$R(\alpha|\mu) = R(\alpha|\mu, \beta)$$

et

$$R(\beta|\mu) = R(\beta|\mu, \alpha).$$

5° Tous les modèles présentés ci-dessus sont dits surparamétrés ou modèles-S; ils sont très utiles pour comprendre comment obtenir les sommes des carrés (notation R).

Par contre, pour écrire les hypothèses [HOCKING, 1985; SEARLE, 1987], on a recours aux modèles dits des moyennes ou modèles-M.

Ainsi, en considérant l'exemple complet du tableau 1 et le modèle (6), il y a douze paramètres à estimer ($\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\gamma}_{11}, \hat{\gamma}_{12}, \hat{\gamma}_{13}, \hat{\gamma}_{21}, \hat{\gamma}_{22}, \hat{\gamma}_{23}$) au départ de six cellules; le modèle (6) est donc de type S. Mais, on peut aussi écrire un autre modèle [SPEED, HOCKING, HACKNEY, 1978], à savoir :

$$y_{ijk} = \mu_{ij} + \epsilon_{ijk}. \quad (7)$$

Dans ce cas, il y a seulement six paramètres à estimer ($\hat{\mu}_{11}, \hat{\mu}_{12}, \hat{\mu}_{13}, \hat{\mu}_{21}, \hat{\mu}_{22}, \hat{\mu}_{23}$); le modèle (7) est un modèle-M.

Le lien entre ces deux modèles peut être établi si :

$$\mu_{ij} = \mu + \alpha_i + \beta_j + \gamma_{ij}.$$

Tableau 2. Tableau d'analyse de la variance.

Sources de variation	Degrés de liberté	Sommes des carrés	Notations R
Types de sols	1	181,5	$R(\alpha \mu)$
Types de sondes	2	49,0	$R(\beta \mu)$
Interaction	2	3,0	$R(\gamma \mu, \alpha, \beta)$
Erreur	18	112,5	
Totaux	23	346,0	
Effet moyen	1	42.336,0	$R(\mu)$
Totaux généraux	24	42.682,0	
Effet moyen et types de sols	2	42.517,5	$R(\mu, \alpha)$
Effet moyen et types de sondes	3	42.385,0	$R(\mu, \beta)$
Effet moyen, types de sols et types de sondes	4	42.566,5	$R(\mu, \alpha, \beta)$
Effet moyen, types de sols, types de sondes et interaction	6	42.569,5	$R(\mu, \alpha, \beta, \gamma)$
Types de sols ajustés par l'effet moyen et les types de sondes	1	181,5	$R(\alpha \mu, \beta)$
Types de sondes ajustés par l'effet moyen et les types de sols	2	49,0	$R(\beta \mu, \alpha)$

3. CAS AVEC CELLULES VIDES

3.1. Introduction

Dans le cas d'un tableau de données avec cellules vides, on décrit les hypothèses les plus courantes pour tester l'effet du premier facteur A (quatre hypothèses) ou effets des sols (paragraphe 3.2), pour tester l'effet du second facteur B (quatre hypothèses) ou effets des sondes (paragraphe 3.3) et l'interaction (une hypothèse) entre ces deux facteurs (paragraphe 3.4), sachant que bien souvent, le chercheur croit tester les hypothèses suivantes :

$$H_0 : \alpha_1 = \dots = \alpha_i = \dots = \alpha_p,$$

$$H_0' : \beta_1 = \dots = \beta_j = \dots = \beta_q$$

et
$$H_0'' : \gamma_{11} = \dots = \gamma_{ij} = \dots = \gamma_{pq} = 0,$$

sans connaître précisément le sens exact à donner aux α_i , β_j et γ_{ij} .

Préalablement, le tableau 3 décrit les modèles M et S conformément aux équations (6) et (7) pour l'exemple avec cellules vides issus du tableau 1; ce dernier tableau reprend également les effectifs des échantillons entre parenthèses (n_{ij}).

Tableau 3. Modèles M et S.

	Sonde 1 (j=1)	Sonde 2 (j=2)	Sonde 3 (j=3)
Sol 1 (i=1)	$\mu_{11}-(2)$ $\mu + \alpha_1 + \beta_1 + \gamma_{11}$	$\mu_{12}-(1)$ $\mu + \alpha_1 + \beta_2 + \gamma_{12}$	$\mu_{13}-(2)$ $\mu + \alpha_1 + \beta_3 + \gamma_{13}$
Sol 2 (i=2)	$\mu_{21}-(2)$ $\mu + \alpha_2 + \beta_1 + \gamma_{21}$	$\mu_{22}-(3)$ $\mu + \alpha_2 + \beta_2 + \gamma_{22}$	- -

3.2. Hypothèses du premier facteur

1° La première hypothèse $H_0^{(1)}$ ou hypothèse de type 1 se rapporte aux moyennes pondérées des niveaux du premier facteur :

$$H_0^{(1)} : \frac{2\mu_{11} + \mu_{12} + 2\mu_{13}}{5} = \frac{2\mu_{21} + 3\mu_{22}}{5}.$$

Si on remplace chaque μ_{ij} par l'expression correspondante du modèle S, l'hypothèse devient [IEMMA, PALM, CLAUSTRIAUX, 1993] :

$$\begin{aligned}
 H_0^{(1)} : & (\alpha_1 - \alpha_2) \\
 & - \frac{2}{5}\beta_2 + \frac{2}{5}\beta_3 \\
 & + \frac{2}{5}\gamma_{11} + \frac{1}{5}\gamma_{12} + \frac{2}{5}\gamma_{13} - \frac{2}{5}\gamma_{21} - \frac{3}{5}\gamma_{22} = 0
 \end{aligned}$$

sachant que la somme des carrés (SC) de $H_0^{(1)}$ correspond à :

$$SC H_0^{(1)} = R(\alpha|\mu).$$

Comme on peut le constater, cette hypothèse est fonction des effectifs des échantillons et elle est très différente de la simple écriture :

$$H_0 : \alpha_1 - \alpha_2 = 0.$$

Dès lors, la différence entre les deux sols est liée, non seulement, aux effets des sondes, mais aussi à l'interaction des deux facteurs, sauf si ces deux dernières sources de variation sont supposées nulles. Signalons dès à présent qu'en pratique, il n'est pas judicieux de tester une hypothèse sur les sols sans avoir préalablement vérifié l'absence d'interaction entre les sols et les sondes.

La figure 2 présente la procédure et les résultats obtenus par SAS pour l'hypothèse $H_0^{(1)}$. Remarquons que ce qui est indiqué sous l'intitulé "Type 1 - Estimable Functions for : A" est la description de $H_0^{(1)}$, si L2 vaut 1 et si A,B et A*B correspondent respectivement à α , β et γ . De plus, $R(\alpha|\mu)$ vaut 52,90. Mais, même s'il apparaît une valeur de Snedecor significative, il n'est pas évident d'expliquer ce que signifie ce rejet d'hypothèse nulle :

$$\alpha_1 + \left(\frac{2}{5}\beta_3 + \frac{2}{5}\gamma_{11} + \frac{1}{5}\gamma_{12} + \frac{2}{5}\gamma_{13} \right) \neq \alpha_2 + \left(\frac{2}{5}\beta_2 + \frac{2}{5}\gamma_{21} + \frac{3}{5}\gamma_{22} \right).$$

Notons encore que le sens à donner aux autres sommes des carrés est expliqué dans les paragraphes suivants. En particulier, il faut remarquer que l'information relative aux sondes (B) présentée dans les tableaux d'analyse de la variance ne correspond pas pour celles-ci à l'hypothèse de type 1, mais à l'hypothèse de type 2 (paragraphe 3.3).

La figure 3 illustre l'application par le logiciel Minitab. Ce qui se réfère à $R(\beta|\mu, \alpha)$ et à $R(\gamma|\mu, \alpha, \beta)$ est en relation avec les paragraphes 3.3 et 3.4.

PROCEDURE

```
PROC GLM;
CLASS A B;
MODEL P205=A B A*B/SS1 E1;
```

RESULTATS

General Linear Models Procedure
Type 1 Estimable Functions for: A

Effect	Coefficients	
INTERCEPT	0	
A	1	L2
	2	-L2
B	1	0
	2	-0.4*L2
	3	0.4*L2
A*B	1 1	0.4*L2
	1 2	0.2*L2
	1 3	0.4
	2 1	-0.4*L2
	2 2	-0.6*L2

Dependent Variable: P205

Source	DF	Sum of Squares	F Value	Pr > F
Model	4	67.43333333	5.06	0.0525
Error	5	16.66666667		
Corrected Total	9	84.10000000		
	R-Square	C.V.	P205 Mean	
	0.801823	4.485852	40.7000000	

Source	DF	Type II SS	F Value	Pr > F	Hypotheses	R(.)
A	1	52.90000000	15.87	0.0105	$H_0^{(1)}$	$R(\alpha \mu)$
B	2	14.48571429	2.17	0.2094	$H_0^{(2)}$	$R(\beta \mu, \alpha)$
A*B	1	0.04761905	0.01	0.9095	$H_0^{(1)}$	$R(\gamma \mu, \alpha, \beta)$

Figure 2. Procédure et résultats SAS pour $H_0^{(1)}$.

PROCEDURE

```
MTB > glm c3=C1 c2 c1*c2
```

RESULTATS

General Linear Model

Factor	Levels	Values		
SOLS	2	1	2	
SONDES	3	1	2	3

Analysis of Variance for P205

Source	Model DF	Reduced DF	Seq SS	Hypothèses	R(.)
SOLS	1	1	52.900	$H_0^{(1)}$	$R(\alpha \mu)$
SONDES	2	2	14.486	$H_0^{(2)}$	$R(\beta \mu, \alpha)$
SOLS*SONDES	2	1+	0.048	$H_0^{(1)}$	$R(\gamma \mu, \alpha, \beta)$
Error	4	5	16.667		
Total	9	9	84.100		

+ Rank deficiency due to empty cells, unbalanced nesting or collinearity.

No storage of results or further analysis will be done.

Figure 3. Procédure et résultats Minitab pour $H_0^{(1)}$.

2° La deuxième hypothèse $H_0^{(2)}$ ou hypothèse de type 2 concerne la comparaison des moyennes des sols pondérées par leurs effectifs et ajustées par les moyennes marginales des sondes également pondérées par leurs effectifs, pour éliminer leurs effets. En pratique, s'il n'y a pas d'effet sol, la moyenne pondérée de l'effet du sol 1 ou du sol 2 doit être égale à la moyenne pondérée de la marge calculée selon les expressions suivantes :

$$\bar{C}_1 = \frac{2\mu_{11} + 2\mu_{21}}{4},$$

$$\bar{C}_2 = \frac{1\mu_{12} + 3\mu_{22}}{4},$$

et
$$\bar{C}_3 = \frac{2\mu_{13}}{2}.$$

En réalité, $H_0^{(2)}$ cache deux hypothèses équivalentes :

$$H_0^{(2.1)} : \frac{2\mu_{11} + \mu_{12} + 2\mu_{13}}{5} = \frac{2\bar{C}_1 + \bar{C}_2 + 2\bar{C}_3}{5}$$

et

$$H_0^{(2.2)} : \frac{2\mu_{21} + 3\mu_{22}}{5} = \frac{2\bar{C}_1 + 3\bar{C}_2}{5}.$$

Si on considère les deux parties gauches des hypothèses $H_0^{(2.1)}$ et $H_0^{(2.2)}$, on retrouve $H_0^{(1)}$.

En développant $H_0^{(2.1)}$ ou $H_0^{(2.2)}$ avec les notations du modèle S, on obtient :

$$H_0^{(2)} : (\alpha_1 - \alpha_2) + \frac{4}{7}\gamma_{11} + \frac{3}{7}\gamma_{12} - \frac{4}{7}\gamma_{21} - \frac{3}{7}\gamma_{22} = 0,$$

sachant que :

$$SC H_0^{(2)} = R(\alpha | \mu, \beta).$$

Si comme prévu les β_j sont éliminés (effets des sondes), il faut surtout remarquer que γ_{13} n'intervient pas et, par conséquent, sous cette hypothèse les observations de la sonde 3 ne sont pas utilisées.

La figure 4 illustre cette hypothèse par l'intermédiaire de SAS. La fonction estimable correspond à $H_0^{(2)}$ si L2 vaut 1 et si A et A*B correspondent à α et γ . Dans ce cas, $R(\alpha | \mu, \beta)$ vaut 26,04.

A nouveau, le sens à donner au rejet de $H_0^{(2)}$ n'est pas évident à interpréter :

$$\alpha_1 + \frac{4}{7}\gamma_{11} + \frac{3}{7}\gamma_{12} \neq \alpha_2 + \frac{4}{7}\gamma_{21} + \frac{3}{7}\gamma_{22}.$$

Minitab ne fournit pas directement pour le type 2 des résultats pour les sols, mais bien pour les sondes (paragraphe 3.3). Il est possible de les obtenir en changeant l'ordre d'entrée des facteurs dans le modèle (paragraphe 3.3; figure 9).

PROCEDURE

```
PROC GLM;
CLASS A B;
MODEL P205=A B A*B/SS2 E2;
```

RESULTATS

General Linear Models Procedure							
Type II Estimable Functions for: A							
Effect	Coefficients						
INTERCEPT	0						
A	1	L2					
	2	-L2					
B	1	0					
	2	0					
	3	0					
A*B	1 1	0.5714*L2					
	1 2	0.4286*L2					
	1 3	0					
	2 1	-0.5714*L2					
	2 2	-0.4286*L2					

Source	DF	Type II SS	F Value	Pr > F	Hypothèses	R(.)
A	1	26.03571429	7.81	0.0382	$H_0^{(2)}$	$R(\alpha \mu, \beta)$
B	2	14.48571429	2.17	0.2094	$H_0^{(2)}$	$R(\beta \mu, \alpha)$
A*B	1	0.04761905	0.01	0.9095	$H_0^{(1)}$	$R(\gamma \mu, \alpha, \beta)$

Figure 4. Procédure et résultats SAS pour $H_0^{(2)}$.

3° La troisième hypothèse $H_0^{(3)}$ ou hypothèse de type 3 se fonde sur la comparaison des moyennes des sols non pondérées par leurs effectifs et ajustées par les moyennes marginales des sondes non pondérées par leurs effectifs. Le principe de son établissement est analogue à $H_0^{(2)}$ en considérant les moyennes non pondérées de la marge :

$$\bar{C}_1 = \frac{\mu_{11} + \mu_{21}}{2},$$

$$\bar{C}_2 = \frac{\mu_{12} + \mu_{22}}{2},$$

et $\bar{C}_3 = \mu_{13},$

PROCEDURE

```
PROC GLM;
CLASS A B;
MODEL P205=A B A*B/SS3 E3;
```

RESULTATS

General Linear Models Procedure
Type III Estimable Functions for: A

Effect	Coefficients	
INTERCEPT	0	
A	1	L2
	2	-L2
B	1	0
	2	0
	3	0
A*B	1 1	0.5*L2
	1 2	0.5*L2
	1 3	0
	2 1	-0.5*L2
	2 2	-0.5*L2
	2 3	0

Source	DF	Type III SS	F Value	Pr > F	Hypotheses	R(.)
A	1	25.19047619	7.56	0.0404	$H_0^{(3)}$	$R(\alpha \mu, \beta, \gamma)$
B	2	13.77543860	2.07	0.2218	$H_0^{(3)}$	$R(\beta \mu, \alpha, \gamma)$
A*B	1	0.04761905	0.01	0.9095	$H_0^{(1)}$	$R(\gamma \mu, \alpha, \beta)$

Figure 5. Procédure et résultats SAS pour $H_0^{(3)}$.

et l'hypothèse est :

$$H_0^{(3)} : \frac{\mu_{11} + \mu_{12} + \mu_{13}}{3} = \frac{\bar{C}_1 + \bar{C}_2 + \bar{C}_3}{3}$$

ou
$$H_0^{(3)} : \frac{\mu_{11} + \mu_{12}}{2} = \frac{\mu_{21} + \mu_{22}}{2} \quad \text{ou} \quad \mu_{11} + \mu_{12} = \mu_{21} + \mu_{22}.$$

Dans le cas particulier où le premier facteur n'a que deux niveaux, ce qui est le cas pour les sols, μ_{13} n'apparaît pas, c'est-à-dire que toute colonne avec une cellule vide n'est pas présente dans l'hypothèse.

Quant à l'hypothèse avec le modèle S, elle est :

$$H_0^{(3)} : (\alpha_1 - \alpha_2) + \frac{1}{2}\gamma_{11} + \frac{1}{2}\gamma_{12} - \frac{1}{2}\gamma_{21} - \frac{1}{2}\gamma_{22} = 0,$$

sachant que $SC H_0^{(3)} = R(\alpha | \mu, \beta, \gamma)$, qui dans l'exemple vaut 25,19. La figure 5 présente les résultats pour le logiciel SAS et non pour le logiciel Minitab puisque celui-ci ne propose pas de résultats dans le cas de l'hypothèse de type 3 avec des cellules vides.

Le rejet de l'hypothèse est plus simple à interpréter :

$$\alpha_1 + \bar{\gamma}_1 \neq \alpha_2 + \bar{\gamma}_2.,$$

si $\bar{\gamma}_1$ et $\bar{\gamma}_2$ sont les moyennes de l'interaction de chaque type de sol avec les sondes.

4° La quatrième hypothèse $H_0^{(4)}$ ou hypothèse de type 4 est aussi une hypothèse sur les moyennes des lignes non pondérées par les colonnes, mais en considérant toujours la dernière ligne du tableau comme ligne de référence. Dès lors, l'hypothèse est fonction de la ligne de référence, c'est-à-dire de la présentation des résultats dans le tableau de données; elle n'est donc pas unique (voir paragraphe 3.3, 1°).

L'hypothèse de type 4 est :

$$H_0^{(4)} : \mu_{11} + \mu_{12} = \mu_{21} + \mu_{22}.$$

Remarquons que dans ce cas particulier où $p=2$, $H_0^{(4)}$ est identique à $H_0^{(3)}$.

La figure 6 illustre cette hypothèse pour SAS. La notation R n'est pas utilisée. Minitab n'envisage pas ce type d'hypothèse.

PROCEDURE

```
PROC GLM;
CLASS A B;
MODEL P205=A B A*B/SS4 E4;
```

RESULTATS

General Linear Models Procedure
Type IV Estimable Functions for: A

Effect		Coefficients
INTERCEPT		0
A	1	L2
	2	-L2
B	1	0
	2	0
	3	0
A*B	1 1	0.5*L2
	1 2	0.5*L2
	1 3	0
	2 1	-0.5*L2
	2 2	-0.5*L2
	2 3	0

NOTE: Other Type IV estimable functions exist.

Source	DF	Type III SS	F Value	Pr > F	Hypotheses	R(.)
A	1*	25.19047619	7.56	0.0404	$H_0^{(4)}$	-
B	2*	6.00000000	0.90	0.4636	$H_0^{(4)}$	-
A*B	1	0.04761905	0.01	0.9095	$H_0^{(1)}$	$R(\gamma \mu, \alpha, \beta)$

* NOTE: Other Type IV Testable Hypotheses exist which may yield different SS.

Figure 6. Procédure et résultats SAS pour $H_0^{(4)}$.

3.3. Hypothèses du deuxième facteur

1° Les hypothèses de type 1 à 4 pour le modèle M dans le cas de l'étude des sondes (deux degrés de liberté) sont reprises ci-dessous, sachant que d'autres hypothèses équivalentes peuvent être écrites.

Type 1

$$H_0^{(1)} : \begin{cases} \frac{2\mu_{11} + 2\mu_{21}}{4} = \frac{\mu_{12} + 3\mu_{22}}{4} \\ \frac{2\mu_{11} + 2\mu_{21}}{4} = \mu_{13} \end{cases}$$

Type 2

$$H_0^{(2)} : \begin{cases} \frac{2\mu_{11} + 2\mu_{21}}{4} = \frac{2\bar{L}_1 + 2\bar{L}_2}{4} \\ \frac{\mu_{12} + 3\mu_{22}}{4} = \frac{\bar{L}_1 + 3\bar{L}_2}{4} \end{cases}$$

où
$$\bar{L}_1 = \frac{2\mu_{11} + \mu_{12} + 2\mu_{13}}{5}$$

$$\bar{L}_2 = \frac{2\mu_{21} + 3\mu_{22}}{5}$$

Type 3

$$H_0^{(3)} : \begin{cases} \frac{\mu_{11} + \mu_{21}}{2} = \frac{\bar{L}_1 + \bar{L}_2}{2} \\ \frac{\mu_{12} + \mu_{22}}{2} = \frac{\bar{L}_1 + \bar{L}_2}{2} \end{cases}$$

où
$$\bar{L}_1 = \frac{\mu_{11} + \mu_{12} + \mu_{13}}{3}$$

$$\bar{L}_2 = \frac{\mu_{21} + \mu_{22}}{2}$$

Type 4

Sur base du tableau initial (tableau 3), en considérant la troisième colonne comme référence, on peut écrire :

$$H_0^{(4)} : \begin{cases} \mu_{11} = \mu_{13} \\ \mu_{12} = \mu_{13} \end{cases}$$

Dans ce cas, les moyennes de la sonde 2 ne sont pas utilisées.

Toujours en se référant à la colonne 3, on peut aussi substituer les moyennes des sondes 2 et 3 et ainsi obtenir une deuxième hypothèse de type 4 :

$$H^{(4)} : \begin{cases} \mu_{11} = \mu_{12} \\ \mu_{21} = \mu_{22} \end{cases}$$

2° Pour le modèle S, les hypothèses des quatre types pour le facteur sondes (B) sont présentées à la figure 7 sous la forme des fonctions estimables SAS.

Pour obtenir l'hypothèse (deux degrés de liberté), il suffit d'égaliser L2 à 1 et L3 à 0 pour obtenir la première ligne (premier degré de liberté) et l'inverse pour avoir la seconde ligne (deuxième degré de liberté).

3° Les résultats obtenus avec les logiciels SAS et Minitab sont présentés aux figures 8 et 9 sur lesquelles sont indiquées les hypothèses et les notations R.

3.4. Hypothèses pour l'interaction

Les hypothèses pour l'interaction comprennent deux aspects à savoir, d'une part, l'hypothèse de l'effet de l'interaction sol x sonde au départ d'un modèle l'incluant et, d'autre part, le cas des hypothèses sur les deux facteurs sols et sondes lorsque le modèle d'ajustement n'inclut pas l'effet de l'interaction.

1° Si on considère un modèle avec l'effet de l'interaction, l'hypothèse générale relative à celle-ci est :

$$H_0^{(i)} : \mu_{ij} + \mu_{i.} - \mu_{.j} + \mu = 0,$$

pour tout i et j, sachant que $\mu_{i.}$ et $\mu_{.j}$ sont les moyennes théoriques pour chaque niveau des deux facteurs.

L'application de cette formule à l'exemple avec cellule vide aboutit à la fonction estimable unique (un degré de liberté) suivante :

$$H_0^{(i)} : \mu_{11} + \mu_{22} - \mu_{12} - \mu_{21} = 0.$$

Effect	Type 1		Type 2		Type 3		Type 4	
		Coefficients		Coefficients		Coefficients		Coefficients
INTERCEPT		0		0		0		0
B	1	L2		L2		L2		L2
	2	L3		L3		L3		L3
	3	-L2-L3		-L2-L3		-L2-L3		-L2-L3
A	1	-0.5*L2-0.75*L3		0		0		0
	2	0.5*L2+0.75*L3		0		0		0
A*B	1	0.5*L2		0.7857*L2+0.4286*L3		0.75*L2+0.25*L3		L2
	1	0.5*L2		0.2143*L2-0.5714*L3		0.25*L2+0.75*L3		L3
	1	-L2-L3		-L2-L3		-L2-L3		-L2-L3
	2	0.5*L2		0.2143*L2-0.4286*L3		0.25*L2-0.25*L3		0
	2	0.75*L3		0.2143*L2-0.4286*L3		-0.25*L2+0.25*L3		0
	2							

Figure 7. Fonctions estimables SAS pour les sondes (B).

PROCEDURES

PROC GLM ;
 CLASS A B ;
 MODEL P205= B A A*B/SS1 SS2 SS3 SS4 E1 E2 E3 E4 ;

RESULTATS

Dependent Variable: P205

Source	DF	Sum of Squares	F Value	Pr > F	Hypotheses	R(.)
Model	4	67.43333333	5.06	0.0525		$R(\alpha, \beta, \gamma \mu)$
Error	5	16.66666667				
Corrected Total	9	84.10000000				
	R-Square	C.V.	P205 Mean			
	0.801823	4.485852	40.7000000			

Source	DF	Type I SS	F Value	Pr > F	H ₀ ⁽¹⁾	R(β μ)
B	2	41.35000000	6.20	0.0442	H ₀ ⁽¹⁾	R(β μ)
A	1	26.03571429	7.81	0.0382	H ₀ ⁽²⁾	R(α μ, β)
A*B	1	0.04761905	0.01	0.9095	H ₀ ⁽¹⁾	R(γ μ, α, β)

Source	DF	Type II SS	F Value	Pr > F	H ₀ ⁽²⁾	R(β μ, α)
B	2	14.48571429	2.17	0.2094	H ₀ ⁽²⁾	R(β μ, α)
A	1	26.03571429	7.81	0.0382	H ₀ ⁽²⁾	R(α μ, β)
A*B	1	0.04761905	0.01	0.9095	H ₀ ⁽¹⁾	R(γ μ, α, β)

Source	DF	Type III SS	F Value	Pr > F	H ₀ ⁽³⁾	R(β μ, α, γ)
B	2	13.77543860	2.07	0.2218	H ₀ ⁽³⁾	R(β μ, α, γ)
A	1	25.19047619	7.56	0.0404	H ₀ ⁽³⁾	R(α μ, β, γ)
A*B	1	0.04761905	0.01	0.9095	H ₀ ⁽¹⁾	R(γ μ, α, β)

Source	DF	Type IV SS	F Value	Pr > F		
B	2*	6.00000000	0.90	0.4636	-	-
A	1*	25.19047619	7.56	0.0404	-	-
A*B	1	0.04761905	0.01	0.9095	H ₀ ⁽¹⁾	R(γ μ, α, β)

* NOTE: Other Type IV Testable Hypotheses exist which may yield different SS.

Figure 8. Procédure et résultats SAS pour les sondes (B).

PROCEDURE

```
MTB > glm c3=C2 c1 c2*c1
```

RESULTATS

General Linear Model

Factor	Levels	Values		
SONDES	3	1	2	3
SOLS	2	1	2	

Analysis of Variance for P205

Source	Model DF	Reduced DF	Seq SS	Hypothèses	R(.)
SONDES	2	2	41.350	$H_0^{(1)}$	$R(\beta \mu)$
SOLS	1	1	26.036	$H_0^{(2)}$	$R(\alpha \mu, \beta)$
SONDES*SOLS	2	1+	0.048	$H_0^{(1)}$	$R(\gamma \mu, \alpha, \beta)$
Error	4	5	16.667		
Total	9	9	84.100		

+ Rank deficiency due to empty cells, unbalanced nesting or collinearity.

Figure 9. Procédure et résultats Minitab pour les sondes (B).

Comme on le constate, μ_{13} n'apparaît pas; dès lors, on ne peut tester qu'une partie de la nullité de l'interaction présente et ce test n'est pas un critère parfait de choix d'un modèle avec ou sans interaction. En conséquence aussi, pour les quatre types présentés antérieurement, la fonction estimable est toujours la même (figure 10).

Quant à la valeur des sommes des carrés relatifs à l'exemple, elle vaut $R(\gamma|\mu, \alpha, \beta) = 0,048$ (figures 2, 3, 4, 5, 6 et 8).

Pour l'exemple, puisque le test associé à cette hypothèse conduit à accepter l'hypothèse nulle ($F_{\text{Obs}} = 0,01$), cela signifie qu'une partie seulement de l'interaction sol \times sonde est nulle ($\mu_{11} + \mu_{22} = \mu_{12} + \mu_{21}$).

En pratique, comme déjà signalé antérieurement, c'est cette hypothèse qu'il convient de tester en premier lieu pour éviter toutes les difficultés d'interprétation soulevées aux paragraphes 3.2 et 3.3 tout en n'oubliant pas la remarque formulée à l'alinéa précédent.

De même, avant de procéder à l'analyse de toutes les données du tableau incomplet, il conviendrait peut-être de se prononcer sur l'intérêt de garder la cellule $i = 1, j = 3$, en réalisant par exemple un contraste particulier à l'issue d'une simple analyse de la variance à un critère de classification (paragraphe 6). Dans ce cas, son élimination, même *a priori*, aboutit à développer ce qui est présenté au paragraphe 4.

Effect	Coefficients		
INTERCEPT			0
A	1		0
	2		0
B	1		0
	2		0
	3		0
A*B	1	1	L7
	1	2	-L7
	1	3	0
	2	1	-L7
	2	2	L7
	2	3	0

Figure 10. Fonction estimable SAS pour $H_0^{(i)}$.

2° Si on considère l'ajustement à un modèle sans interaction, on peut distinguer deux hypothèses différentes de celles définies antérieurement, à savoir, d'une part, l'hypothèse de type 1 et, d'autre part, les hypothèses de types 2, 3 et 4 qui se confondent; l'une quelconque de celles-ci permet de tester parfaitement l'effet des sols ou des sondes ($H_0 : \alpha_1 = \alpha_2$ ou $H_0 : \beta_1 = \beta_2 = \beta_3$).

Evidemment, à la somme des carrés résiduelle correspondent notamment six degrés de liberté.

Les figures 11 et 12 présentent les procédures, les deux fonctions estimables et les tableaux d'analyse de la variance pour l'étude des deux facteurs avec le logiciel SAS; les résultats par le logiciel Minitab sont à la figure 13.

PROCEDURES

```
PROC GLM DATA=SONDE ;
CLASS A B ;
MODEL P205= A B/SS1 SS2 SS3 SS4 E1 E2 E3 E4 ;
```

RESULTATS

Effect	Type I		Types II, III, IV	
	Coefficients		Coefficients	
INTERCEPT	0		0	
A	1	L2	1	L2
	2	-L2	2	-L2
B	1	0	1	0
	2	-0.4*L2	2	0
	3	0.4*L2	3	0

Dependent Variable: P205

Source	DF	Sum of Squares	F Value	Pr > F	Hypothèses	R(.)
Model	3	67.38571429	8.06	0.0158		$R(\alpha, \beta \mu)$
Error	6	16.71428571				
Corrected Total	9	84.10000000				
		R-Square	C.V.	P205 Mean		
		0.801823	4.485852	40.7000000		

Source	DF	Type I SS	F Value	Pr > F	H ₀ ⁽¹⁾	R(α μ)
A	1	52.90000000	18.99	0.0048		
B	2	14.48571429	2.60	0.1537	H ₀ ⁽²⁾	R(β μ, α)

Source	DF	Type II SS	F Value	Pr > F	H ₀ ⁽²⁾	R(α μ, β)
A	1	26.03571429	9.35	0.0223		
B	2	14.48571429	2.60	0.1537	H ₀ ⁽²⁾	R(β μ, α)

Source	DF	Type III SS	F Value	Pr > F	H ₀ ⁽³⁾	R(α μ, β)
A	1	26.03571429	9.35	0.0223		
B	2	14.48571429	2.60	0.1537	H ₀ ⁽³⁾	R(β μ, α)

Source	DF	Type IV SS	F Value	Pr > F	H ₀ ⁽⁴⁾	R(α μ, β)
A	1	26.03571429	9.35	0.0223		
B	2	14.48571429	2.60	0.1537	H ₀ ⁽⁴⁾	R(β μ, α)

Figure 11. Procédure et résultats SAS pour le facteur A (modèle sans interaction).

PROCEDURES

```
PROC GLM DATA=SONDE ;
CLASS A B ;
MODEL P205= B A/SS1 SS2 SS3 SS4 E1 E2 E3 E4 ;
```

RESULTATS

Effect		Type I	Types II, III, IV			
		Coefficients	Coefficients			
INTERCEPT		0	0			
B	1	L2	2	L2		
	2	L3	2	L3		
	3	-L2-L3	3	-L2-L3		
A	1	-0.5*L2-0.75*L3	1			
	2	0.5*L2+0.75*L3	2			

Source	DF	Sum of Squares	F Value	Pr > F	Hypotheses	R(.)
Model	3	67.38571429	8.06	0.0158		$R(\alpha, \beta \mu)$
Error	6	16.71428571				
Corrected Total	9	84.10000000				

	R-Square	C.V.	P205 Mean
	0.801823	4.485852	40.7000000

Source	DF	Type I SS	F Value	Pr > F		
B	2	41.35000000	7.42	0.0239	$H_0^{(1)}$	$R(\beta \mu)$
A	1	26.03571429	9.35	0.0223	$H_0^{(2)}$	$R(\alpha, \mu \beta)$

Source	DF	Type II SS	F Value	Pr > F		
B	2	14.48571429	2.60	0.1537	$H_0^{(2)}$	$R(\beta \mu, \alpha)$
A	1	26.03571429	9.35	0.0223	$H_0^{(2)}$	$R(\alpha \mu, \beta)$

Source	DF	Type III SS	F Value	Pr > F		
B	2	14.48571429	2.60	0.1537	$H_0^{(3)}$	$R(\beta \mu, \alpha)$
A	1	26.03571429	9.35	0.0223	$H_0^{(3)}$	$R(\alpha \mu, \beta)$

Source	DF	Type IV SS	F Value	Pr > F		
B	2	14.48571429	2.60	0.1537	$H_0^{(4)}$	$R(\beta \mu, \alpha)$
A	1	26.03571429	9.35	0.0223	$H_0^{(4)}$	$R(\alpha \mu, \beta)$

Figure 12. Procédure et résultats SAS pour le facteur B (modèle sans interaction)

PROCEDURE

MTB > glm c3=C1 C2

RESULTATS

General Linear Model

Factor	Levels	Values		
SOLS	2	1	2	
SONDES	3	1	2	3

Analysis of Variance for P205

Source	DF	Seq SS	Hypotheses	R(.)	Adj SS	Hypotheses	R(.)	Adj MS
SOLS	1	52.900	$H_0^{(1)}$	$R(\alpha \mu)$	26.036	$H_0^{(3)}$	$R(\alpha \mu, \beta)$	26.036
SONDES	2	14.486	$H_0^{(2)}$	$R(\beta \mu, \alpha)$	14.486	$H_0^{(3)}$	$R(\beta \mu, \alpha)$	7.243
Error	6	16.714			16.714			2.786
Total	9	84.100						

PROCEDURE

MTB > glm c3=C2 C1

RESULTATS

General Linear Model

Factor	Levels	Values		
SONDES	3	1	2	3
SOLS	2	1	2	

Analysis of Variance for P205

Source	DF	Seq SS	Hypotheses	R(.)	Adj SS	Hypotheses	R(.)	Adj MS
SONDES	2	41.350	$H_0^{(1)}$	$R(\beta \mu)$	14.486	$H_0^{(3)}$	$R(\beta \mu, \alpha)$	7.243
SOLS	1	26.036	$H_0^{(2)}$	$R(\alpha \mu, \beta)$	26.036	$H_0^{(3)}$	$R(\alpha \mu, \beta)$	26.036
Error	6	16.714			16.714			2.786
Total	9	84.100						

Figure 13. Résultats Minitab : modèles sans interaction.

4. CAS NON ÉQUIRÉPÉTÉ

4.1. Introduction

Le cas non équirépété se simplifie dans la mesure où $H_0^{(3)} = H_0^{(4)}$ et qu'il n'y a aucun problème pour tester complètement la nullité de l'interaction.

Dès lors, trois hypothèses peuvent être considérées pour étudier chaque facteur.

Sur base des principes énoncés au paragraphe 3, on analyse ce cas de façon synthétique (paragraphe 4.2) et on l'illustre à partir des données du tableau 1 en ne considérant pas les données mises en évidence en italique (paragraphe 4.3).

4.2. Etude des hypothèses

Les hypothèses des types 1, 2 et 3 pour le modèle M sont formulées ci-dessous uniquement pour l'étude du facteur A (sol) en identifiant par μ_{ij} les moyennes théoriques de chaque cellule.

$$H_0^{(1)} : \frac{4\mu_{11} + 3\mu_{12} + 3\mu_{13}}{10} = \frac{4\mu_{21} + 4\mu_{22} + 4\mu_{23}}{12}$$

$$H_0^{(2)} : 2\mu_{11} + \frac{12}{7}\mu_{12} + \frac{12}{7}\mu_{13} - 2\mu_{21} - \frac{12}{7}\mu_{22} - \frac{12}{7}\mu_{23} = 0$$

$$H_0^{(3)} : \frac{\mu_{11} + \mu_{12} + \mu_{13}}{3} = \frac{\mu_{21} + \mu_{22} + \mu_{23}}{3}$$

Les hypothèses pour le modèle S (fonctions estimables) sont présentées à la figure 14 pour chacun des deux facteurs A et B.

Facteur A

Effect	Type I		Type II	Types III et IV
	Coefficients		Coefficients	Coefficients
INTERCEPT		0	0	0
A	1	L2	L2	L2
	2	-L2	-L2	-L2
B	1	0.0667*L2	0	0
	2	-0.0333*L2	0	0
	3	-0.0333*L2	0	0
A*B	1 1	0.4*L2	0.3684*L2	0.3333*L2
	1 2	0.3*L2	0.3158*L2	0.3333*L2
	1 3	0.3*L2	0.3158*L2	0.3333*L2
	2 1	-0.3333*L2	-0.3684*L2	-0.3333*L2
	2 2	-0.3333*L2	-0.3158*L2	-0.3333*L2
	2 3	-0.3333*L2	-0.3158*L2	-0.3333*L2

Facteur B

Effect	Type I		Type II	Types III et IV
	Coefficients		Coefficients	Coefficients
INTERCEPT		0	0	0
B	1	L2	L2	L2
	2	L3	L3	L3
	3	-L2-L3	-L2-L3	-L2-L3
A	1	0.0714*L2	0	0
	2	-0.0714*L2	0	0
A*B	1 1	0.5*L2	0.4737*L2	0.5*L2
	1 2	0.4286*L3	-0.0226*L2+0.4286*L3	0.5*L2
	1 3	-0.4286*L2-0.4286*L3	-0.4511*L2-0.4286*L3	0.5*L2
	2 1	0.5*L2	0.5263*L2	0.5*L2
	2 2	0.5714*L3	0.0226*L2+0.5714*L3	0.5*L2
	2 3	-0.5714*L2-0.5714*L3	-0.5489*L2-0.5714*L3	0.5*L2

Figure 14. Fonctions estimables SAS pour A et B.

4.3. Résultats

Les résultats sont proposés à la figure 15 globalement pour l'ajustement selon la procédure SAS :

$$\text{MODEL Y = A B A*B/etc.}$$

et uniquement pour le type 1 par la procédure :

$$\text{MODEL Y = B A A*B/etc.}$$

L'interprétation synthétique de ceux-ci conduit à constater qu'il y a absence d'interaction sol x sonde ($R(\gamma|\mu, \alpha, \beta) = 3,6992$ et $F_{\text{obs}} = 0,30$).

Dans ce cas, pour tester correctement l'effet des sols ou des sondes, il convient d'utiliser la somme des carrés de type 2 : $R(\gamma|\mu, \alpha, \beta)$ ou $R(\beta|\mu, \alpha)$; elles valent respectivement 138,5865 et 51,5508, c'est-à-dire celles qui sont obtenues par l'ajustement du modèle :

$$\text{MODEL Y = A B /etc,}$$

qui dans ce cas, sont identiques pour les types 2, 3 et 4 (figure 16).

Ce dernier ajustement est utile pour obtenir la somme des carrés à considérer pour l'inférence (101,1992 avec 18 degrés de liberté), valeur qui aurait aussi pu être obtenue (figure 15) en ajoutant $R(\gamma|\mu, \alpha, \beta)$ avec deux degrés de liberté à la somme des carrés de l'erreur (97,5 avec 16 degrés de liberté).

Y = A B A*B

Source	DF	Sum of Squares	F Value	Pr > F	Hypothèses	R(.)
Model	5	205.59090909	6.75	0.0015		$R(\alpha, \beta, \gamma \mu)$
Error	16	97.50000000				
Corrected Total	21	303.09090909				
		R-Square	C.V.	P205 Mean		
		0.678314	5.928837	41.6363636		

Source	DF	Type I SS	F Value	Pr > F		
A	1	150.34090909	24.67	0.0001	$H_0^{(1)}$	$R(\alpha \mu)$
B	2	51.55075188	4.23	0.0335	$H_0^{(2)}$	$R(\beta \mu, \alpha)$
A*B	2	3.69924812	0.30	0.7424	$H_0^{(2)}$	

Source	DF	Type II SS	F Value	Pr > F		
A	1	138.58646617	22.74	0.0002	$H_0^{(2)}$	$R(\alpha \mu, \beta)$
B	2	51.55075188	4.23	0.0335	$H_0^{(2)}$	$R(\beta \mu, \alpha)$
A*B	2	3.69924812	0.30	0.7424	$H_0^{(1)}$	

Source	DF	Type III SS	F Value	Pr > F		
A	1	135.00000000	22.15	0.0002	$H_0^{(3)}$	$R(\alpha \mu, \beta, \gamma)$
B	2	53.18796992	4.36	0.0307	$H_0^{(3)}$	$R(\beta \mu, \alpha, \gamma)$
A*B	2	3.69924812	0.30	0.7424	$H_0^{(1)}$	

Source	DF	Type IV SS	F Value	Pr > F		
A	1	135.00000000	22.15	0.0002	$H_0^{(4)}$	$R(\alpha \mu, \beta, \gamma)$
B	2	53.18796992	4.36	0.0307	$H_0^{(4)}$	$R(\beta \mu, \alpha, \gamma)$
A*B	2	3.69924812	0.30	0.7424	$H_0^{(1)}$	$R(\gamma \mu, \alpha, \beta)$

Y = B A A*B

Source	DF	Type I SS	F Value	Pr > F		
B	2	63.30519481	5.19	0.0183	$H_0^{(1)}$	$R(\beta \mu)$
A	1	138.58646617	22.74	0.0002	$H_0^{(2)}$	$R(\alpha \mu, \beta)$
A*B	2	3.69924812	0.30	0.7424	$H_0^{(1)}$	$R(\gamma \mu, \alpha, \beta)$

Figure 15. Résultats SAS pour le cas non équilibré (modèles avec interaction).

Y - A B

Source	DF	Sum of Squares	F Value	Pr > F	Hypothèses	R(.)
Model	3	201.89166097	11.97	0.0002		$R(\alpha, \beta \mu)$
Error	18	101.19924812				
Corrected Total	21	303.09090909				

R-Square C.V. P205 Mean
 0.666109 5.694815 41.6363636

Source	DF	Type I SS	F Value	Pr > F		
A	1	150.34090909	26.74	0.0001	$H_0^{(1)}$	$R(\alpha \mu)$
B	2	51.55075188	4.58	0.0246	$H_0^{(2)}$	$R(\beta \mu, \alpha)$

Source	DF	Type II SS	F Value	Pr > F		
A	1	138.58646617	24.65	0.0001	$H_0^{(2)}$	$R(\alpha \mu, \beta)$
B	2	51.55075188	4.58	0.0246	$H_0^{(2)}$	$R(\beta \mu, \alpha)$

Source	DF	Type III SS	F Value	Pr > F		
A	1	138.58646617	24.65	0.0001	$H_0^{(3)}$	$R(\alpha \mu, \beta)$
B	2	51.55075188	4.58	0.0246	$H_0^{(3)}$	$R(\beta \mu, \alpha)$

Source	DF	Type IV SS	F Value	Pr > F		
A	1	138.58646617	24.65	0.0001	$H_0^{(4)}$	$R(\alpha \mu, \beta)$
B	2	51.55075188	4.58	0.0246	$H_0^{(4)}$	$R(\beta \mu, \alpha)$

Y - B A

Source	DF	Type I SS	F Value	Pr > F		
B	2	63.30519481	5.63	0.0126	$H_0^{(1)}$	$R(\beta \mu)$
A	1	138.58646617	24.65	0.0001	$H_0^{(2)}$	$R(\alpha \mu, \beta)$

Source	DF	Type II SS	F Value	Pr > F		
B	2	51.55075188	4.58	0.0246	$H_0^{(2)}$	$R(\beta \mu, \alpha)$
A	1	138.58646617	24.65	0.0001	$H_0^{(2)}$	$R(\alpha \mu, \beta)$

Source	DF	Type III SS	F Value	Pr > F		
B	2	51.55075188	4.58	0.0246	$H_0^{(3)}$	$R(\beta \mu, \alpha)$
A	1	138.58646617	24.65	0.0001	$H_0^{(3)}$	$R(\alpha \mu, \beta)$

Source	DF	Type IV SS	F Value	Pr > F		
B	2	51.55075188	4.58	0.0246	$H_0^{(4)}$	$R(\beta \mu, \alpha)$
A	1	138.58646617	24.65	0.0001	$H_0^{(4)}$	$R(\alpha \mu, \beta)$

Figure 16 Résultats SAS pour le cas non équilibré (modèles sans interaction).

5. CAS ÉQUIRÉPÉTÉ

5.1. Introduction

Le cas équirépété est encore plus simple car les hypothèses des quatre types se confondent.

L'illustration de cette affirmation est présentée au départ de la structure du tableau 1 complet (paragraphe 5.2) et de ses données (paragraphe 5.3).

5.2. Etude des hypothèses

Les hypothèses relatives au facteur sol sur base du modèle M et pour les quatre types sont :

$$H_0^{(1)} : \frac{4\mu_{11} + 4\mu_{12} + 4\mu_{13}}{12} = \frac{4\mu_{21} + 4\mu_{22} + 4\mu_{23}}{12}$$

$$H_0^{(2)} : 4\mu_{11} + 4\mu_{12} + 4\mu_{13} - 4\mu_{21} - 4\mu_{22} - 4\mu_{23} = 0$$

$$H_0^{(3)} \text{ et } H_0^{(4)} : \frac{\mu_{11} + \mu_{12} + \mu_{13}}{3} = \frac{\mu_{21} + \mu_{22} + \mu_{23}}{3};$$

toutes les hypothèses peuvent s'écrire :

$$H_0 : \mu_{11} + \mu_{12} + \mu_{13} = \mu_{21} + \mu_{22} + \mu_{23}.$$

Pour le modèle S, l'hypothèse est :

$$H_0 : (\alpha_1 - \alpha_2) + \frac{1}{3}(\gamma_{11} + \gamma_{12} + \gamma_{13}) - \frac{1}{3}(\gamma_{21} + \gamma_{22} + \gamma_{23}) = 0.$$

Les fonctions estimables pour les deux facteurs et l'interaction sont reprises à la figure 17.

5.3. Résultats

Les résultats du cas équirépété sont présentés à la figure 18 pour le modèle avec et sans interaction, type 1 uniquement.

Dans cet exemple, l'hypothèse de l'interaction nulle est acceptée; par contre, il n'en est pas de même pour les hypothèses sur les deux facteurs ($F_{\text{obs}} = 31,43$ pour les sols et $F_{\text{obs}} = 4,24$ pour les sondes).

Facteur A			Facteur B			Interaction		
Effect	Coefficients	Effect	Coefficients	Effect	Coefficients			
INTERCEPT	0	INTERCEPT	0	INTERCEPT	0			
A	1 L2 2 -L2	B	1 L2 2 L3 3 -L2-L3	B	1 0 2 0 3 0			
B	1 0 2 0 3 0	A	1 0 2 0	A	1 0 2 0			
A*B	1 1 0.3333*L2 1 2 0.3333*L2 1 3 0.3333*L2 2 1 -0.3333*L2 2 2 -0.3333*L2 2 3 -0.3333*L2	A*B	1 1 0.5*L2 1 2 0.5*L3 1 3 -0.5*L2-0.5*L3 2 1 0.5*L2 2 2 0.5*L3 2 3 -0.5*L2-0.5*L3	A*B	1 1 L7 1 2 L8 1 3 -L7-L8 2 1 -L7 2 2 -L8 2 3 L7+L8			

Figure 17. Fonctions estimables SAS : cas équiréparté.

Modèle avec interaction

Source	DF	Sum of Squares	F Value	Pr > F
Model	5	233.50000000	7.47	0.0006
Error	18	112.50000000		
Corrected Total	23	346.00000000		

R-Square	C.V.	Y Mean
0.674855	5.952381	42.0000000

Source	DF	Type I SS	F Value	Pr > F
A	1	181.50000000	29.04	0.0001
B	2	49.00000000	3.92	0.0386
A*B	2	3.00000000	0.24	0.7891

Source	DF	Type II SS	F Value	Pr > F
A	1	181.50000000	29.04	0.0001
B	2	49.00000000	3.92	0.0386
A*B	2	3.00000000	0.24	0.7891

Source	DF	Type III SS	F Value	Pr > F
A	1	181.50000000	29.04	0.0001
B	2	49.00000000	3.92	0.0386
A*B	2	3.00000000	0.24	0.7891

Source	DF	Type IV SS	F Value	Pr > F
A	1	181.50000000	29.04	0.0001
B	2	49.00000000	3.92	0.0386
A*B	2	3.00000000	0.24	0.7891

Modèle sans interaction

Source	DF	Sum of Squares	F Value	Pr > F
Model	3	230.50000000	13.30	0.0001
Error	20	115.50000000		
Corrected Tot	23	346.00000000		

R-Square	C.V.	Y Mean
0.666185	5.721721	42.0000000

Source	DF	Type I SS	F Value	Pr > F
A	1	181.50000000	31.43	0.0001
B	2	49.00000000	4.24	0.0291

Figure 18. Résultats SAS pour le cas équiréparté.

6. EN GUISE DE CONCLUSION

1° Les fonctions estimables, notamment celles fournies par SAS, sont très utiles pour écrire les hypothèses du modèle surparamétré et pour comprendre le sens à donner à chaque hypothèse, souvent simplifiée en pratique.

2° En présence de cellules vides, il n'est pas possible de tester l'interaction dans son ensemble, dont le nombre de degrés de liberté est $(p-1)(q-1)-v$, où v est le nombre de cellules vides.

3° Il faut aussi remarquer que toutes les observations récoltées ne sont pas nécessairement utilisées en fonction de l'hypothèse choisie, notamment celle sur l'interaction. Celle-ci devrait donc être parfaitement définie au moment de la conception du plan expérimental, c'est-à-dire bien avant la collecte des observations et l'analyse des résultats ce qui, il faut bien le constater, est une situation peu fréquente en pratique.

4° Cette publication montre encore la complexité des hypothèses et de l'interprétation des résultats, surtout en présence de cellules vides. Cependant, il faut se rappeler qu'il y a toujours une solution plus simple; celle-ci consiste à considérer l'analyse de la variance à un critère de classification, modèle fixe, en fusionnant les facteurs étudiés en une seule source de variation ou de choisir un modèle à deux critères sans interaction, surtout si l'interaction complète ou partielle n'est pas significative (paragraphe 3.4).

7. REMERCIEMENTS

Au terme de cette note, nous exprimons nos remerciements à Madame J. AUSTRAET, qui a largement contribué à sa mise au point.

7. RÉFÉRENCES BIBLIOGRAPHIQUES

- AZAIS J.M. [1994]. Analyse de la variance non orthogonale. L'exemple de SAS/GLM. *Rev. Stat. Appl.*, 42, 27-41.
- BURDICK D.S., HERR D.G. [1980]. Counterexamples in unbalanced two way analysis of variance. *Commun. Statist. Theor. Meth.*, A9 (2), 231-241.
- CLAUSTRIAUX J.J., IEMMA A.F. [1999]. A propos des qualificatifs complet, orthogonal et équilibré en analyse de la variance. *Notes Stat. Inform. (Gembloux)* 99/2, 14 p.
- DAGNELIE P. [1998]. *Statistique théorique et appliquée. Tome 2 : Inférence statistique à une et à deux dimensions*. Bruxelles, De Boeck , 659 p.
- DODGE Y. [1985]. *Analysis of experiments with missing data*. New York, Wiley, 499 p.
- FREUND R.J. [1980]. The case of missing cell. *The Amer. Statist.*, 34, 94-98.
- HERR D.G. [1986]. On the history of anova in unbalanced, Factorial Designs : The first 30 years. *The Amer. Statist.*, 40, 265-270.
- HOCKING R.R. [1985]. *The analysis of linear models*. Monterrey. Brooks/Cole, 385 p.
- IEMMA A.F., PALM R., CLAUSTRIAUX J.J. [1993]. Sobre a construção de Projtores Ortogonais. *Rev. Mat. Estat.* 11, 133-142.
- MURRAY L.W. [1986]. Estimation of missing cells in randomized block and latin square designs. *The Amer. Statist.*, 40, 289-293.
- SEARLE S.R. [1987]. *Linear models for unbalanced data*. New York, Wiley, 536 p.
- SPEED F.M., HOCKING R.R., HACKNEY O.P. [1978]. Methods of analysis of linear models with unbalanced data. *Jour. Amer. Statist. Assoc.*, 73, 105-112.
- URQUART N.S., WEEKS D.L. [1978]. Linear models in messy data : some problems and alternatives. *Biometrics*, 34, 696-705.
- YATES F. [1934]. The analysis of multiple classifications with unequal numbers in the different classes. *Jour. Amer. Statist. Assoc.*, 29, 51-66.
- X [1990]. SAS User's guide : statistics version 6 edition. Cary, SAS Institute, 846 p.
- X [1996]. Minitab reference manual. Enterprise Drive, State College.